

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**



PATENT ABSTRACTS OF JAPAN

(11) Publication number: 10282986 A

(43) Date of publication of application: 23 . 10 . 98

(51) Int. Cl

G10L 3/00
G10L 3/00
G10L 9/10

(21) Application number: 09086486

(22) Date of filing: 04 . 04 . 97

(71) Applicant: HITACHI LTD

(72) Inventor: NAKAGAWA TOMOHITO
MAEJIMA HIDEO

(54) SPEECH RECOGNITION METHOD AND MODEL DESIGN METHOD THEREFOR

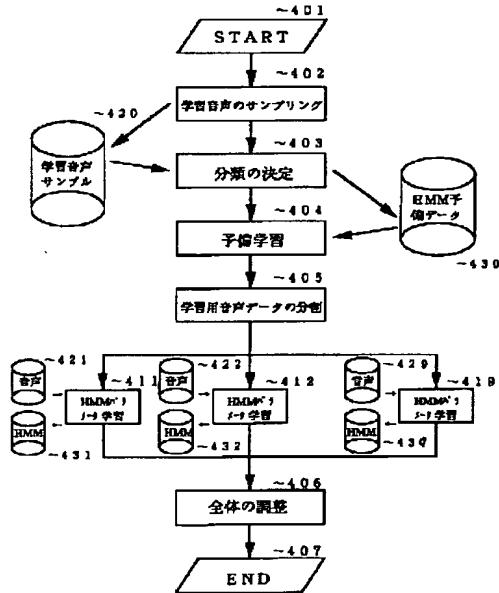
(57) Abstract:

PROBLEM TO BE SOLVED: To realize an adaptive speech recognition system by speaker classification and to easily obtain high performance speech recognition by using fuzzy inference, then estimating the output probability of a hidden Markov model(HMM) from information of similarities of respective classifications and the output probability in respective classifications.

SOLUTION: First of all, speech samples to be learned are collected (S 402). Then, the classifications of the speech samples are decided (A 403). The classifications of the speech sample are decided heuristically, or the optimum classifications are decided from the characteristics of the speech data 420 to be learned. Then, the speech data are answered to respective classifications according to the decided classifications to be divided (S 405). Then, HMM parameters are learned (S 411-419) by using the speech data 421-429 at every classification. At this time, values of preliminary learning are used as state transition probabilities of respective HMMs. Then, adjustment for constituting a whole system is performed from the HMMs learned at every classification (S 406). Methods using the fuzzy

theory and the method using a neural network, etc., are exemplified therefor.

COPYRIGHT: (C)1998,JPO



(51) Int.Cl. ⁸	識別記号	F I
G 1 0 L 3/00	5 3 5	G 1 0 L 3/00
	5 3 1	5 3 5
9/10	3 0 1	5 3 1 K
		9/10
		3 0 1 C

審査請求 未請求 請求項の数 9 OL (全 15 頁)

(21)出願番号 特願平9-86486

(22)出願日 平成9年(1997)4月4日

(71)出願人 000005108
 株式会社日立製作所
 東京都千代田区神田駿河台四丁目6番地

(72)発明者 中川 智仁
 東京都小平市上水本町五丁目20番1号 株式会社日立製作所半導体事業部内

(72)発明者 前島 英雄
 東京都小平市上水本町五丁目20番1号 株式会社日立製作所半導体事業部内

(74)代理人 弁理士 磯村 雅俊

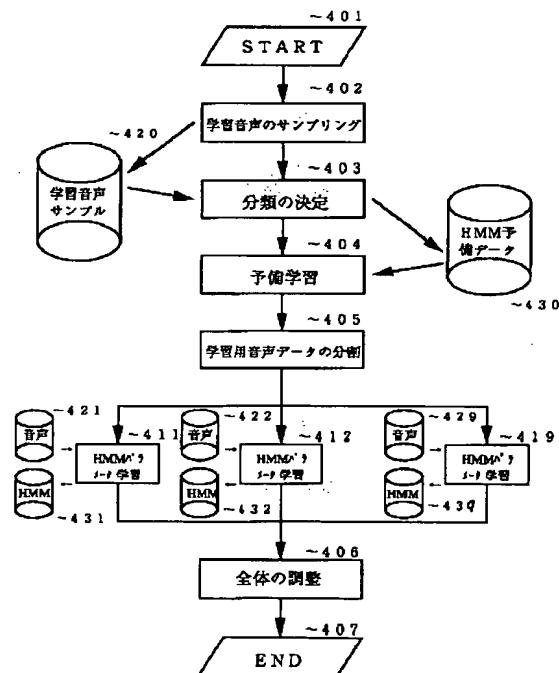
(54)【発明の名称】 音声認識方法およびそのモデル設計方法

(57)【要約】

【課題】高性能の音声認識が可能な話者適応型の音声認識システムを、マイクロコンピュータで効率的に実現する。

【解決手段】老若男女等の話者の特質毎に分類した音声サンプルを用いて、カテゴリ毎に最適な出力関数を決定し、その出力関数を用いて全体の出力関数を決定し、各分類ごとのHMMの出力確率と、話者の属性確率から、話者の属性に合わせた音声認識を実現する。また、この話者分類による認識を推定スコアの算出に用いて、高性能な音声認識を実現する。さらに、この出力確率の計算および推定スコアの算出を、曖昧推論を用い、かつ領域区分法・変数分離法によって高速化する。

実施例1の処理方法(全体図)



【特許請求の範囲】

【請求項1】隠れマルコフモデル（HMM）を用いた音声認識方法のうち、HMMの出力確率を確率密度関数（出力関数）で定義する連續分布型HMMを用いた音声認識システムのモデル設計方法であって、老若男女等の話者の特質ごとにカテゴリーに分類したサンプルを用いて学習し、各カテゴリーごとに最適な出力関数を決定する第1のステップと、該第1のステップで決定された各カテゴリーごとの出力関数を用いて、全体の出力関数を決定する第2のステップとを有することを特徴とする音声認識システムのHMMモデル設計方法。

【請求項2】請求項1に記載の音声認識システムのHMMモデル設計方法において、

前記第1のステップと第2のステップの間に、各カテゴリーごとのHMMの出力確率を曖昧推論で記述するためには、まず単一のガウス分布に相当する曖昧推論系を構成した後、そのデータを用いて混合ガウス分布系の学習パラメータを計算して曖昧規則系を構築し、計算に際しては、領域区分法に従って操作領域を決定した後、変数分離法に従って一変数テーブルを作成する第3のステップを設けたことを特徴とする音声認識システムのHMMモデル設計方法。

【請求項3】請求項1に記載の音声認識システムのHMMモデル設計方法において、

前記第2のステップは、HMMの出力確率を計算し、適当な評価関数を定義して該関数が最大または最小にするパラメータを評価し、該パラメータに対し評価を良くする方向に修正した後、収束判定を行う各処理からなることを特徴とするHMMモデル設計方法。

【請求項4】請求項1に記載の音声認識システムのHMMモデル設計方法において、

前記第3のステップは、ニューラルネットワークを用いており、入力された学習データに基づいて誤差逆伝播法により結合係数の調整を行い、パラメータを修正した後、収束判定を行うことを特徴とするHMMモデル設計方法。

【請求項5】HMMを用いた音声認識方法のうち、HMMの出力確率を出力関数で定義する連續分布型HMMの出力確率計算方法において、

上記各HMMの状態ごとに、各カテゴリーごとの出力関数を決定するパラメータを備え、入力された音声の分析情報から、話者のカテゴリーをその特質で明確に推定する第1のステップと、

該第1のステップで推定したカテゴリーの情報と音声の分析情報から、出力確率を決定する第2のステップとを有することを特徴とする連續分布型HMMの出力確率計算方法。

【請求項6】HMMを用いた音声認識方法のうち、HMMの出力確率を出力関数で定義する連續分布型HMMを

用いた音声認識方法において、

上記各HMMの状態ごとに、各カテゴリーごとの出力関数を決定するパラメータを備え、入力された音声の分析情報から、話者のカテゴリーをモデルに修正された重み係数を付して確率的に推定する第1のステップと、該第1のステップで確率的に推定したカテゴリーの情報と音声の分析情報から、出力確率を決定する第2のステップとを有することを特徴とする連續分布型HMMの出力確率計算方法。

10 【請求項7】請求項5または6に記載の連續分布型HMMの出力確率計算方法において、

前記第1のステップは、音声データである特徴ベクトルをニューラルネットワークの入力層に入力し、該ニューラルネットワークの各ノードごとに計算を実行し、出力に相当するノードの計算値を最終出力とすることを特徴とする連續分布型HMMの出力確率計算方法。

【請求項8】請求項5または6に記載の連續分布型HMMの出力確率計算方法において、

前記第2のステップは、ビタビのベストファーストサーチの原理を用い、まずHMMパラメータを読み込み、入力音声の特徴（特徴ベクトル）と上記HMMパラメータを用いて探索を行った後に、スコアを評価して最適性を判定し、最適候補とならないときには次の探索点に処理を移しながら同じ処理を繰り返し、残った候補を最適パスとして、その対応するモデルを認識候補として出力することを特徴とする連續分布型HMMの出力確率計算方法。

【請求項9】HMMを用いた音声認識方法のうち、HMMの出力確率を出力関数で定義する連續分布型HMMを

30 用いた音声認識方法において、各HMMの状態ごとに、各カテゴリーごとの出力関数を決定するパラメータを備え、入力された音声の分析情報から、話者のカテゴリーを確率的に推定する第1のステップと、確率的に推定したカテゴリーの情報と音声の分析情報から、話者とカテゴリーの類似性を評価する第2のステップと、

評価した類似性に基づいて出力確率を適応させる第3のステップとを有することを特徴とする話者適応方法。

【発明の詳細な説明】

40 【0001】

【発明の属する技術分野】本発明は、隠れマルコフモデル（以下、HMM）（ここでは連續分布型HMM）を用いた音声認識を、マイクロコンピュータ（以下、マイコン）のような簡単な処理装置で効率的に実行する音声認識方法およびその方法を用いたモデル設計方法に関する。

【0002】

【従来の技術】

（1）従来の音声認識装置について

50 従来より、音声認識装置101は、例えば図1に示すよ

うなシステム構成で実現される。マイクなどの音声入力装置102で入力された音声信号は、AD変換装置103でデジタル信号に変換される。特徴抽出装置104では、一定の区間（以下、フレーム）を定常的と看做した音の性質を分析する。認識装置105では、フレームごとに抽出されたパラメータの時系列的な変動過程を言葉ごとに比較し、最も近いと思われる言葉を認識結果として出力する（以下、この比較を距離計算と呼ぶ）。

（2）従来のHMM型音声認識について

（2.1）HMM型音声認識の基本原理について、HMM型音声認識では、前述の距離計算を確率的に評価するため、図2に示すようなマルコフモデルを考える。ここで、マルコフモデルとは、マルコフ性（将来の状態が現在の状態によってのみ規定される）の仮定された確率的な状態遷移モデルのことである。この確率的な挙動は、状態間の遷移確率によって表現される。ここで、状態とは『あ』、『い』等の音源と考えればよく、人によって音源の内容が異なったり、音の特徴がずれたりする。HMMでは、認識する音源をモデルで表現しようとするもので、『あ』の音源に対するモデル、『い』の音源に対するモデル、・・・をそれぞれ作成する。そして、それが具体的に音の特徴、例えば周波数分布（いくつかのパラメータで表現される）がどう変化するかで認識する。それぞれの状態については、出力する確率が与えら*

$$bs(xt) = \sum W_{ks} \frac{1}{\Pi(2\pi\sigma^2_{ki})^{1/2}} \exp\left(-\sum \frac{(x_i - \mu_{ki})^2}{2\pi\sigma^2_{ki}}\right)$$

これは、ガウス分布の線形結合で表現されるため、出力確率分布は基底となるガウス分布の数が多ければ多いほど精密になる。しかし、計算量もそれに比例する。この方式では、計算量が多く必要となるが、認識精度は高いとされている。一方、離散分布型HMMは、特徴ベクトルを量子化し、その各々に対して出力確率分布を定義する。これは、図3（b）に示すような階段状の確率密度関数を与えたものと等価である。離散分布型HMMでは、個々の確率計算の計算負荷が連続分布型に比べて小さくなる。しかし、量子化された特徴ベクトルごとに出力確率のテーブルが必要となり、また、量子化における誤差が認識精度を劣化させるとと言われている。

【0004】（3）

（3.1）従来の話者適応化について、代表的な話者適応方式としては、最大事後確率推定法（MAP推定法）や移動ベクトル場平滑化法（VFS法）あるいはこれらの複合法（逐次型MAP/VFS法）などがある（例えば、高橋他：『逐次型話者適応方式MAP/VFSにおける分散適応』、音響学会講演論文集、2-5-5（1995）参照）。ここで、話者適応化とは、人によって声の高低や声の広がり等あり、パターンに違いがあるので、これらはHMMにより統計的に吸収できるが、理想的モデルにするために、標準のモデルから離れた人に対してモデル

*れており、それぞれがどの程度の出力であるかにより表現される。実際に認識するときには、音の性質がどのように変化するかを計算して、最も確率の高いものが最もらしいとして認識される。このように、音声認識では、状態ごとに、ある特徴ベクトルを出力する確率と状態遷移確率とが与えられている。つまり、音声認識は、特徴ベクトルの関数で表現される。HMMは、各認識要素（単語や音素）ごとに一連の状態によって表現される。例えば、図2に示すような left-to-right 型のHMM

10 が数多く採用されている。そして、認識処理においては、各認識要素のモデルごとに、与えられた特徴ベクトル列（以下、観測系列）を出力しうる確率（尤度）を計算し、最も尤度の高いモデルを認識候補として選択する。

【0003】（2.2）HMMにおける出力確率の表現について、前述のHMMは、出力確率の表現形態の違いにより、連続分布型HMMと離散分布型HMMとに類別できる。連続分布型HMMは、各状態における出力確率を図3（a）に示すように連続分布で表現する。このため、観測された特徴ベクトルは、量子化されることなく出力確率が計算される。一般に、確率密度関数の表現には、混合ガウス分布が用いられる。この混合ガウス分布は、下式（数1）で記載される。

【数1】

$$\mu_{MAP} = \frac{n}{n+r} m + \frac{r}{n+r} \mu_0$$

これは、ガウス分布の線形結合で表現されるため、出力確率分布は基底となるガウス分布の数が多ければ多いほど精密になる。しかし、計算量もそれに比例する。この方式では、計算量が多く必要となるが、認識精度は高いとされている。一方、離散分布型HMMは、特徴ベクトルを量子化し、その各々に対して出力確率分布を定義する。これは、図3（b）に示すような階段状の確率密度関数を与えたものと等価である。離散分布型HMMでは、個々の確率計算の計算負荷が連続分布型に比べて小さくなる。しかし、量子化された特徴ベクトルごとに出力確率のテーブルが必要となり、また、量子化における誤差が認識精度を劣化させるとと言われている。

30 3.1.1）話者適応の具体的方法について、上述のMAP推定法においては、「事前分布に依存する事後確率を最大にする」ようにパラメータを推定する。具体的には下式（数2）のように修正する。

【数2】

$$\mu_{MAP} = \frac{n}{n+r} m + \frac{r}{n+r} \mu_0$$

ただし、nはサンプル数、mはサンプル平均、rはデータの確からしさ

また、上述のVFS法は、写像法の一種で、「話者の特徴=標準者の特徴+個人差ベクトル」と考え、少数のサンプルの平均ベクトルを個人差ベクトルと看做すことにより適応させる。具体的には下式（数3）のように修正する。

【数3】

$$Vp = C^R p - C^I p$$

ただし、 $C^I p$ は初期モデルのガウス分布、 $C^R p$ は再学習後のガウス分布

【0005】(3.2) 従来の統計的話者分類について
統計的な話者分類を用いる方法は、多くの場合、話者適応化を含めて議論されている。離散分布型HMMにおいて、話者カテゴリーごとに作製したHMMをマージする方法（例えば、今村：『統計的話者分類による話者適応HMM音声認識』、信学技報 SP91-17 (1991) 参照）、あるいは、混合ガウス分布型HMMにおいて、話者カテゴリーごとに作製したHMMをマージ（重み係数を修正する）する方法（例えば、『話者混合逐次分割法による不特定話者音声認識と話者適応』、信学論 A Vol. J77-A, No. 2 pp. 103-111 (1994) 参照）などが提案されている。統計的話者分類を用いる場合、例えば、男と女とでは統計的にパターンの分布が明確に分かれているので、この特性を用いて簡単に分類することができる。前述の混合ガウス分布型HMMについて述べると、前式（数1）のガウス分布に重みをかけて加算することにより、正規分布を幾つか重ねたものが形成される。これは、幾つかのピークを有するパターンであり、話者適応化によりそれぞれの特徴があるパターン全体をそっくりずらしてもよいのであるが、複数のピークが均等にずれている場合には、処理が複雑になるので、話者の特性として、男であることがわかつていれば、男の特徴的パターンを使って分類すれば簡単な処理で分類することができる。

(3.3) 従来の連続分布型HMMの高速計算法について
混合ガウス分布の高速計算法は、例えばスカラ量子化（例えば、S. Sagayama : ON THE USE OF SCALAR QUANTIZATION FOR, ICASSP95, I-pp. 213-217 (1995) 参照）などの方法が効率的である。また、本願の発明者らは、出力確率計算に曖昧推論を用いた音声認識法を本願出願前に提案している。しかし、この方法では、写像法がテーブル化されているので、使い難いという問題がある。

【0006】

【発明が解決しようとする課題】連続分布型HMM音声認識を対象にする場合、FVS法やMAP推定法のように、出力関数を直接的に修正する方法では、変数分離法による曖昧推論を用いた方法やスカラ量子化のようなテーブル化手法は使内できない。また、混合分布の重み係数を修正する方法では、テーブル化手法は使えるものの、混合ガウス分布に限られるし、また前式（数1）では、 $\text{addlog} [\log ((\exp(a) + \exp(b)))]$ の計算が必要になるため、計算負荷が多大になる。すなわち、話者適応化方法の処理では、補正をかけて標準モデルに平行移動するので、モデル自体が変化してしまい、テーブル内容を計算し直さなければならなくなる。本発明の目的は、このような従来の課題を解決し、離散分布型HMMまたは混合分布型HMM以外の音声認識システムでも、話者分類による適応的音声認識システムが実現でき、領域区分法・変数分離法等のテーブル化手法の導入が可能で、高性能な音声認識が比較的容易に実現できるマイクロコンピュータ向けの音声認識方法およびそれを用いたモデル

設計方法を提供することにある。

【0007】

【課題を解決するための手段】上記目的を達成するため、本発明の音声認識方法では、連続分布型HMMのうち、混合ガウス分布を用いるのではなく、出力確率計算に曖昧推論を用いる方法（いわゆる半連続型HMM）を想定している。曖昧推論は、領域区分法・変数分離法

（例えば、中川他：『制御用マイコン向け高速ファジイ推論方式』、情報処理学会第43回全国大会講演論文

10 集、I-101参照）を用いることにより高速に実行できるが、事前にテーブルにするため、話者適応化が難しい。そこで本発明では、統計的話者分類による方法を用いる。すなわち、音韻的特徴の類似した各分類ごとに、曖昧推論により出力関数を定義して、各分類との類似性の情報と、各分類における出力確率（曖昧推論によって得られる）から、曖昧推論を用いて、HMMの出力確率を推定する。音声認識では、2つのポイントがあり、その1つは認識処理の方法であり、他の1つはどのように認識システムを構成するかという認識処理のための基礎となる学習に関する方法である。つまり、後者は、認識では出力確率のための関数を決定することが大きな問題であって、具体的な学習のテーマとその学習したモデルから認識システムを生成することである。以下の実施例では、主として認識システムを生成するための学習に関する例と、音声認識の方法（出力確率）に関する例とを挙げる。

【0008】

【発明の実施の形態】以下、本発明の実施例を、図面により詳細に説明する。図4は、本発明の実施例1を示す

30 音声認識システムの動作フローチャートである。ここでは、本発明で示した音声認識システムを、実際に設計する手順を示している。処理が開始されると（ステップ401）、まず最初に学習すべき音声サンプルを収集する（ステップ402）。HMMによる不特定話者認識においては、さまざまなタイプの音声を学習させる必要がある。これらは、HMMの構成単位（例えば、音節）ごとに、発話者の属性（性別・年齢等）とともに管理する。次に、音声サンプルの分類を決定する（ステップ403）。これらの音声サンプルの分類は、ヒューリスティックに決定する（例えば、老若男女別とか）か、あるいは学習すべき音声データ420の特徴から最適な分類を決定する。次に、音声データ420からHMMを予備学習する（ステップ404）。これは、例えば通常の混合ガウス分布型連続HMM音声認識用の学習処理を用いて、混合数を上記の分類数と等しくさせるように学習させる。そして、学習の結果を、HMM予備データ430に格納する。

40 【0009】次に、ステップ403で決定した分類に従って、音声データ420を各分類に対応させ区分する（ステップ405）。例えば、発話者の属性に応じてヒ

ヒューリスティックに分類したならば、その属性に従って分類すれば良い。これらは、その分類に従って音声データ421～429に分割して格納する。ヒューリスティックの分類とは、例えば男と女、子供と成人と老人等に分類することである。次に、各分類ごとに音声データ421～429を用いて、HMMパラメータを学習する（ステップ411～419）。この時、それぞれのHMMの状態遷移確率は、ステップ404で行った予備学習の値を用いる。これは、各HMMの状態遷移確率を等しくするためである。次に、各分類ごとに学習したHMMより、全体のシステムを構成するための調整を行う（ステップ406）。これは、曖昧理論を用いる方法やニューラルネットワークを用いる方法などが考えられる（この例については、他の実施例として後述する）。あるいは、単純な実現方法としては、各分類の属性確率を「重み」とみなし、一種の加重平均を計算することで、個々のシステムと全体とを関連付けることも可能である。ステップ406で、各個別のシステムと全体との相互関連を学習させることにより、システムが完成する（ステップ407）。

【0010】図5は、図4におけるHMMパラメータ学習（ステップ411）の詳細処理のフローチャートであり、図6は、曖昧推論による出力確率の推定の説明図である。なお、図4におけるステップ412～419についても同様である。ここでは、図6に示すように、HMMの出力確率が曖昧推論で記述される例を示す。曖昧推論は、経験的なノウハウを、図6(a)のような曖昧推論によって記述する。曖昧規則は、入力変数と出力変数の関係を曖昧に（曖昧変数で）記述している。曖昧変数は、具体的には図6(b)のようなメンバシップ関数で定義される。この曖昧規則は、一般にヒューリスティックな方法によって（試行錯誤で）決定される。勿論、そのようにしても十分実現できる。もっとも、本実施例では、図5に示すように、一度混合ガウス分布系の学習パラメータを計算してから、曖昧規則系を再構築する方法を示す。間接的再学習でも可能であるが、処理が簡単であるため上記の方法を採用した。図5のステップ501～507は、全体として図4のステップ411のサブルーチン相当する。処理が開始されたサブルーチンでは（ステップ501）、まず最初に、音声データ421を用いて、混合ガウス分布系のパラメータを学習する（ステップ502）。これは、一般的な Baum-Welch のアルゴリズムを用いれば求められる。学習した混合ガウス分布系のパラメータは、データ512に格納する。

【0011】次に、ステップ502で学習したパラメータを用いて、曖昧推論系を構成する（ステップ503、504）。なお、この方法の詳細は、発明者らが既に提案している方法である。すなわち、まず単一のガウス分布に相当する曖昧推論系を構成する（ステップ503）。この結果をデータ511に格納する。次に、この

データ511のデータ、つまり各構成ガウス分布の推論系を用いて、混合ガウス分布系に相当する曖昧推論系を構成する（ステップ504）。この曖昧推論系の計算は、通常の方法（例えば、マムダニの方法や代数積一加算一重心法）では、計算時間が多大になるので、領域区分法および変数分離法によって高速化する。そのため、領域区分法に従って操作領域の決定を行った後（ステップ505）、変数分離法に従って一変数のテーブル作成を行う（ステップ506）。テーブル値は、ファイル431に格納される。なお、領域区分法・変数分離法は、特に入力数の多大な曖昧推論を高速に実行する方法である（例えば、中川他：『制御用マイコン向き高速ファジイ推論方式』情報処理学会第43回全国大会講演論文集I, pp. 101-102 (1991) 参照）。図6(a)に示すように記述された出力確率は、代数積一加算一重心法では、下式（数4）のように表現される。

【数4】

$$P_0 = \frac{\sum_{i=1}^m \omega_i(x_1, x_2, \dots) \times k_{mi}}{\sum_{i=1}^m \omega_i(x_1, x_2, \dots) \times k_{ai}}$$

$$k_{mi} = \int_E \{p \times \mu_i(p)\} dp$$

$$k_{ai} = \int_E \{\mu_i(p)\} dp$$

領域区分法において、適用領域・操作領域・操作規則は次式（数5）のように定義される。

【数5】

曖昧規則 i の適用領域 D_i は

$$D_i = \{(x_1, x_2, \dots) | \omega_i(x_1, x_2, \dots) > 0\}$$

操作領域 D^k ($k=1, 2, \dots, m$) は、

$$\forall (x_1, x_2, \dots) \in D^k, \forall i \in \bar{M}^k, \omega_i(x_1, x_2, \dots) = 0$$

かつ

$$\bigcap_k (D^k) = \emptyset, \bigcup_k (D^k) = X$$

操作領域 M^k ($k=1, 2, \dots, m$) は、

$$M^k = \{i | \omega_i(x_1, x_2, \dots) > 0, \exists (x_1, x_2, \dots) \in D^k\}$$

この時、出力確率は、下式（数6）のように操作領域ごとに与えられる。

【数6】

$$P_0 = \frac{\sum_{i \in M^k} \omega_i(x_1, x_2, \dots) \times k_{mi}}{\sum_{i \in M^k} \omega_i(x_1, x_2, \dots) \times k_{ai}} \quad ((x_1, x_2, \dots) \in D^k)$$

変数分離法では、領域区分法において適合度を限界積で評価する。これにより、出力確率は、操作領域ごとに定義された以下の推論式（数7）で計算できる。

【数7】

$$P_0 = \frac{Fx1_\alpha(x1) + Fx2_\alpha(x2) + \dots + K_\alpha}{Fx1_\beta(x1) + Fx2_\beta(x2) + \dots + K_\beta} \quad [(x1, x2, \dots) \in D^k]$$

$$Fx1_\alpha(x1) = \sum_{i \in M^k} \mu_i x_{1i}(x1) \times k_{mi}$$

$$Fx1_\beta(x1) = \sum_{i \in M^k} \mu_i x_{1i}(x1) \times k_{ai}$$

$$Fx2_\alpha(x2) = \sum_{i \in M^k} \mu_i x_{2i}(x2) \times k_{mi}$$

$$Fx2_\beta(x2) = \sum_{i \in M^k} \mu_i x_{2i}(x2) \times k_{ai}$$

$$K_\alpha = n(M^k) \times (1 - l_p) \times k_{mi}$$

$$K_\beta = n(M^k) \times (1 - l_p) \times k_{ai}$$

ただし、 l_p は入力変数の数
 $n(A)$ は集合Aの要素数

図5のステップ505は、言い替えれば前式(数4)を与えることに他ならない。また、ステップ506は、具体的には、前式(数7)中の定数および一変数の関数を事前計算し、テーブル化する処理である。

【0012】図7は、図4における全体の調整処理(ステップ406)の詳細動作フローチャートであり、図8は、ニューラルネットワークの構成およびそれを用いた調整処理のフローチャートである。ここでは、各分類の出力確率を用いて、全体のシステムを下式(数8)のように記述する。

【数8】

$$P = \sum \alpha_k P_k \quad \begin{array}{l} \text{ただし、} \alpha_k \text{は分類属性値} \\ P_k \text{は分類} k \text{における出力確率} \end{array}$$

要するに、ここでは各分類属性値を線形的に補正し、これを「重み係数」として出力確率を計算するものである。ここでは、原理的に一番単純なため、実施例にした。なお、後述の実施例2および実施例3では他の例を示す。図7のステップ701～707は、図4のステップ406のサブルーチンに相当する。まず、出力確率を計算する(ステップ702)。ここでは、前式(数8)に従って出力確率を計算する。ただし、初回は適当な初期値(例えば、すべて1)が与えられているものとし、それに基づいて計算される。次に、パラメータを評価する(ステップ703)。具体的には、ある適当な評価関数を定義し、それを最大(あるいは最小)にすることを考える。あるいは、一定のサンプルを入力し、その認識性能を評価指標にしても良い。本実施例では、適当な評価関数を定義することを想定する。次に、評価を良くする方向にパラメータを修正する(ステップ704)。具体的には、評価関数をパラメータに対して偏微分(数値的には差分)し、その方向(勾配方向)にパラメータを(一定の小さい幅で)修正する。次に、収束判定を行う(ステップ705)。ここでは、前の値と現在の値が十分小さい一定値(ϵ)より小さくなることをもって、収束と看做す。収束しない場合には、ステップ702へ戻る。収束した場合には、補正係数を確定し(ステップ706)、処理を終了する(ステップ707)。

【0013】次に、実施例2を説明する。実施例2は、実施例1における図4の全体調整処理(ステップ406)を他の方法で行う例である。ここでは、図8に示す

10

ようなニューラルネットワーク(以下、NN)を用いている。NNは、図8(a)のようなニューロン素子をネットワーク状に結合したもの(図8(b)参照)である。ニューロン素子の出力は、一般に入力の総和に対する関数として表現される。例えば、下式(数9)に示すような、シグモイド関数で表現されることが多い。

【数9】

$$f(x) = \frac{1}{1 + \exp(-x)}$$

10 これを、実システムに適用する場合、例えば図8(b)のような階層型ネットワークを用いることができる。階層型ネットワークでは、左から入力層・中間層・出力層となっており、本実施例では、入力層の各ノードと分類を対応させ(具体的には、分類の属性値を入力にする)、出力層に重み係数を対応させている。そして、所望の入出力関係が得られるように、結合係数を調整する。結合係数の調整は、一般的な誤差逆伝播法(BP法)によって実行できる。これは、下式(数10)で示される。

20 【数10】

$$\Delta \omega_{ji}(n) = -\eta \frac{\partial E}{\partial \omega_{ji}(n)} + \alpha \Delta \omega_{ji}(n-1)$$

ただし、 η 、 α は学習係数

上式(数10)では、パラメータを修正する(丁度、誤差が逆方向=出力層→入力層に伝わっていくイメージがあるため誤差逆伝播法と呼ばれる)。この計算を、収束するまで繰り返し実行する。これを具体的に表現すると、図8(c)のようになる。図8のステップ601～605は図4のステップ406のサブルーチンとなる。

30 入力された入出力関係(学習データ)に基づいて(ステップ602)、前式(数10)に従ってパラメータを修正する(ステップ603)。そして、収束判定を行い(ステップ604)、収束しない場合にはステップ602へ戻り、収束した場合にはステップ605へ進む。

【0014】収束判定(ステップ604)においては、NNの場合、最小にする評価関数(通常、エネルギー関数と呼ばれる)は、多嶺性(極小値が複数存在する)を示すため(何故ならば、極小値=最小値とは限らない)、図7で示したような方法(線形システムの収束判定)は適用できない(局所最適の回避問題)。多くの場合、局所最適を回避する方法として、確率的な方法—例えば、シミュレーテッドアニーリング(SA)法などが用いられる。SA法は、一定の確率頻度で、勾配を逆方向に探索する。この確率頻度は、温度関数によって制御される。SA法では、この温度関数を徐々に小さく(すなわち、確率頻度を少しづつ小さくする)することで、確率的に準最適な解を求めようとするものである。このSA法を用いた場合、収束条件は、温度関数がTが0に極めて近づいた時に、パラメータが収束を満たすことが必要になる。このように、NNでは処理が複雑にはなる

40 50

が、大枠では、図8 (c) の手順で実行できる。

【0015】図9は、本発明の実施例3を示す全体構成図である。図9には、本発明による話者適応型音声認識システムの実現例の概要が示されている。処理が開始されると(ステップ801)、音声入力装置(マイク)810から音声が入力される。この音声は、A/D変換され(あるいは、A/D変換器を用いて)デジタル信号としてシステムに取り込まれる(ステップ802)。次に、音声の特徴分析を行う(ステップ803)。これは、Levinson-Durbin法によってLPC(Liner Predictive Coding)系のパラメータを求める方法や、FFT(高速フーリエ変換)によって周波数スペクトルを求める方法がある。ステップ804~806は、本発明の特有な処理となる。すなわちステップ803で用いた分析結果(特徴パラメータ)を用いて、入力音声の属する話者分類を決定する(ステップ804)。

【0016】次に、この結果を用いて話者適応化を行う(ステップ805)。これに基づき、認識処理を行う(ステップ806)。計算結果は、出力結果ファイル812へ出力する。上記ステップ801~807のうち、ステップ804以外は、実施例1および実施例2で求めたパラメータを用いて計算できる。しかし、話者属性の評価は、認識処理の時に用わなければならない。この詳細について、図10に示す。図10は、図9における話者類類の分析処理(ステップ804)の詳細説明図である。ここでは、実施例2と同様の階層型NNを用いた例を示す。ここでは、入力として各特徴ベクトル、出力として各分類ごとの話者属性を対応させる。NNの結合係数の学習は、実施例2で示した手順と同様である。この場合、出力は図9(d)を用いて計算できる。まず処理が開始されると(ステップ906)、音声データ(ここでは、特徴ベクトル)が入力層に入力される(ステップ907)。次に、各ノードごとに、下式(数11)に従って計算が実行される(ステップ908)。

【数11】

$$z_o = \sum_m \omega_{mo} y_m$$

$$y_m = \sum_i \omega_{im} x_i$$

ただし、 z_o は、出力ノードの値
 y_m は、中間ノードの値
 x_i は、入力ノードの値

ここで、出力に相当するノードの計算値が出力となる。この出力に相当するノードの計算の終了を待って終了する(ステップ909)。なお、結合係数の学習方法のフローについては、図8(c)のフローと同じであるため説明を省略する。

【0017】図11は、本発明の実施例4を示す音声認識処理の全体フローチャートである。また、図12は、図11のフローで必要となるベストファーストサーチとスコア推定の説明図である。実施例4では、本発明を認

識処理における探索処理のスコア推定に適用した例を示したものである。スコア推定は、出力計算を行って最も大きな出力を求めるため、フレームに関してビタビサーチを行って経路を探索し、時間が1歩進む毎にどのような状態をとるかを決定し、最も大きなパスを選択する。

そのときに、最も大きなパスを選択すれば計算量は減少するが、最適パスの欠落をなくすためにベストファーストサーチによりある程度の推定を行う。これは、探索していないが、見込みで評価して推定する。図11のステップ1001~1007のうち、ステップ1002(認識照合)を除いて実施例3の処理と共通している。本実施例では、認識照合においてビタビのベストファーストサーチを用いた例を示す。ベストファーストサーチでは、他のサーチには存在しない「スコア推定」という処理が必要になる。そして、このスコア推定の方法(具体的には、スコア関数の設定)によって認識性能が大きく左右される。このスコア推定のためには、システムデータ811の他に、スコア推定に関する情報1010が必要になる。

【0018】本実施例を説明する前に、簡単にベストファーストサーチとスコア関数について説明する。音声認識では、認識要素ごとに定義されたHMMのうち、入力音声の特徴(正確には、特徴ベクトルの列:以下、観測ベクトル)を出力する確率の最も高いHMMを選択する。しかし、「隠れマルコフ」の名の通り、入力音声からは状態遷移系列は分からず。そこで、どのように該当するHMMを探し出すかが問題になる。これを、探索(サーチ)と呼ぶこととする。サーチで、一番単純な方法は図12(a)に示すように、全てのパス(状態遷移の経路)をサーチする方法である。横軸に時間(t)を示し、縦軸に状態を示している。これは、動的計画法を用いて簡単に実現できる。これに対して、ベストファーストサーチは、図12(b)に示すように、各フレームごとに最も確率の高いパスを選択して、それ以外を落とす方法である。単純なインプリメントでは、最大あるいは上位何候補かを残すだけで処理を行う場合もあるが、この場合、最適パスを切る可能性があり、認識性能は劣化する。従って、最適パスをできるだけ落とさないようにパスを選択しなくてはならない。

【0019】いわゆる、「最適性の保証」が重要な問題になる。これは、例えばサーチの途中の段階でも最後までのパスのスコアが分かれれば(実際は未知であるが)、トータルで最適なパスを判定することができる。そこで、未だ処理の済んでいない部分のスコアを予測し、それを加えた上で最適性を判定することが必要になる。ここで、未知の部分のスコアを評価することをスコア推定と呼ぶ。ここでは、このスコア推定において、(できるだけ)最適性を保証するよう推定する(スコア関数を設定する)ことが問題となっている。スコア関数が一定の条件(A*条件)を満たす時、最適性が保証される(A*探

索）。これを、図12(c)に示す。いま、ノード(i, j, n)からノード(i+1, k, n)までのスコアをs((i, i+1), (j→k), n)のように定義する。このとき認識問題は、s((0, 1), (0, Jn), n)が最大となるパス(最適パス)を決定する問題となる。いま、図12(c)のように、探索が中間点Cまで進行したとする。このとき、全体のスコアは、既に探索が終了した評価(G)と終了していない部分の評価(H)の和で表現できる。すなわち、 $F(i, j, n) = G(i, j, n) + H(i, j, n)$ のようになる。ノードCにおけるスコア関数の推定値は、 $f(i, j, n) = g(i, j, n) + h(i, j, n)$ で推定する。この推定スコアが、 $h(i, j, n) > H(i, j, n)$ を満たす場合、A*探索となり最適解が保証される。しかし、厳密に「A*条件」を満たすようにスコア関数を設定すると、計算負荷が大きくなることが多い。そこで、厳密に条件を満たさなくても、近似的に的確なスコア関数を設定し、認識性能を維持しつつ計算量を削減することが求められる。

【0020】そこで、実施例4では、実施例1および実施例2の方法を用いて、推定スコアを設定する方法、およびその推定スコアを用いて認識照合(図11のステップ1002)を実現する方法を説明する。ここでは、スコア推定のためのHMM(以下、推定HMM)を用いる方法を例にして述べる。スコア推定HMMは、図12(d)のように、逆方向の状態遷移によって定義される。これは、終状態からのスコアを計算するためのものである。このHMMの学習は、通常のHMMと全く同様に実行できる。ただし、入力する特徴ベクトルは逆にする。例えば、フレーム0～20のデータなら、通常は0, 1, … の順になるが、ここでは、20, 19, … の順になる。従って、本発明の方法(実施例1、実施例2)の方法も、全く同じように適用できる。すなわち、特徴ベクトルを入れる順番を逆にするだけで良い。実施例4では、スコア推定においても話者分類による話者適応化を行う。そのため、推定HMMおよびスコア関数は各話者分類ごとに設定し、スコア推定時において、話者属性に応じてスコアの適応化(以下、スコア関数の話者適応化)を行う。

【0021】図13は、図11で示した実施例4の処理に必要なスコア情報(データ1010)の作成処理のフローチャートである。これは、具体的には、スコア推定のためのデータ(以下、スコアデータ)である。実施例4では、推定HMMを用い、さらにこれを曖昧理論によって簡略化する方法を示す。処理が開始されると(ステップ1201)、まず推定HMMの学習を行う(ステップ1202)。これは、図5に示した実施例1の方法で実現できる。勿論、推定HMMは、各話者分類ごとに定義する。この、推定HMMを用いてスコア関数を設定し、設定されたスコア関数のデータ値をデータファイル1210に格納する。次に、ステップ1202で決定したスコア関数を表現する曖昧推論系を構成する(ステッ

プ1203)。曖昧推論系は、図6に示したような形式で表現される。本実施例では、ヒューリスティックに決定することを想定する。尚、スコア関数より曖昧推論系を決定する他の例は、実施例5で説明する。このスコア関数の曖昧推論系は、データファイル1211に格納する。本実施例では、先の実施例で示した領域区分法・変数分離法を用いるため、操作領域を決定した後(ステップ1204)、テーブルを作成する(ステップ1205)。これは、先に示した方法と同じである。このテーブルのデータ値をデータファイル1010に格納する。このテーブルにより、スコア推定が可能になる。

【0022】図14は、図11の実施例4の認識照合処理(ステップ1002)の詳細処理のフローチャートである。ここで、ステップ1301～1309は、図11のステップ1002のサブルーチンに相当する。処理が開始されると(ステップ1301)、システムデータ811のHMMパラメータを読み込む(ステップ1302)。ここで、観測ベクトルは、図11のステップ803で既に分析されている。また、初期設定も同時にを行う。例えば、初期探索点として、現在位置を、初期時刻(t=0)初期状態(s=0)に設定する。これは、図12(b)のトrellis(状態を縦軸に、時間を横軸にとった2次元空間)上で表現すれば、点Aに相当する。ステップ1303以降では、観測ベクトルと、HMMパラメータを用いて探索を行う。図11(b)において、初期探索点Aから状態遷移できるノードは、B, Cである。そこで、まずBに探索点を設定する(ステップ1303)。次に、スコアを評価し(ステップ1304)、最適性を判定する(ステップ1305)。ここで、最適候補となり得れば登録し(ステップ1306)、最適候補とならなければ次の探索点(例えばC)に処理を移す(ステップ1303)。この処理で、最適候補は、B, Cのいずれかに決定されるので、次にここで決定したノードを現在位置として、同様の処理を継続する。これを最終状態に到達するまで繰り返す(ステップ1307)。最終状態と判定したならば(ステップ1307)、残った候補を最適パスとして(ステップ1308)、その対応するモデルを認識候補として出力する。出力結果は、ファイル812に格納される。

【0023】図15は、本発明の実施例5を示す推定HMMから曖昧推論系を構成する動作フローチャートである。実施例4の曖昧推論系の決定処理(図13のステップ1203)は、ヒューリスティックな手法を用いているのに対して、実施例5の同じ処理(図15のステップ1401)では、A*条件を満たすように決定する方法を示している。ステップ1401以外は、図13のフローの動作と全く同じであるため、説明を省略する。図16は、図15における曖昧推論系の決定処理(ステップ1401)の詳細を示すフローチャートである。ここでは、曖昧規則はヒューリスティックに決定し、メンバシ

ップ関数を反復計算によって最適に調整する。図16のステップ1501～1509は、図15のステップ1401のサブルーチンに相当する。図16に示すように、処理が開始されると（ステップ1501）、初期設定が行われる（ステップ1502）。ここでは、データファイル1210によって与えられたスコア関数を入力し、また、このデータに基づいて曖昧規則を決定する。これは、スコア関数の代表点を表現するように与えると良い。次に、メンバシップ関数を調整（修正）する（ステップ1503）。ここでは、例えばモンテカルロ的な確率的手法が適用できる。すなわち、修正の候補を確率的に選択し、修正が条件を満たした場合、遡って更新される（条件を満たさない場合修正しない）。ここで修正したメンバシップ関数に対して、出力を評価する（ステップ1504）。次に、A*条件を満たすか否かを判定し（ステップ1505）、満たさなければステップ1503からやり直す。また、A*条件を満たせば、誤差評価を行う（ステップ1506）。ここでは、例えばデータ1210で与えられたスコア関数との二乗誤差を評価関数にすれば良い。次に、この誤差二乗関数が小さくなつた場合のみ修正を行い（ステップ1507）、それ以外の場合、ステップ1503で行った修正を無効にする（もとに戻す）。このような試行を一定回数繰り返すことにより、最適解に近い解（準最適解）が求められる（ステップ1508）。最適解は、データファイル1211に格納される。

【0024】

【発明の効果】以上説明したように、本発明によれば、離散分布型HMMあるいは混合分布型HMM以外の音声認識システムでも、話者分類による適応的音声認識システムが実現できる。特に、出力確率を曖昧推論系で定義した場合、高速化手法として領域区分法・変数分離法などのテーブル化手法の導入が必須になるが、このような方法を用いても、話者適応型音声認識が実現できるようになるため、高性能の音声認識が比較的容易に実現可能になる。これまで話者分類型適応方式においてのスコア関数の議論はなされていなかったが、本発明によれば、音声認識システムの高速化のためには、ベストファーストサーチのような高速探索法の導入が必須になるので、話者分類型適応方式のスコア関数も高速に推定可能になる。

【図面の簡単な説明】

【図1】従来における音声認識装置の例を示す構成図である。

【図2】従来におけるleft-to-right型HMMの構成例を示す図である。

【図3】HMMの出力確率の表現形態を示す図である。

【図4】本発明の実施例1を示す音声認識システム構成法の動作フローチャート（全体図）である。

【図5】図4の実施例1におけるパラメータ学習（ステップ411）の処理フローチャートである。

【図6】曖昧推論による出力確率の推定を示す表現例図である。

【図7】図4の実施例1における全体調整（ステップ406）の処理フローチャートである。

【図8】本発明の実施例2を示す全体調整（ステップ406）の処理フローチャートおよびニューラルネットワークの説明図である。

【図9】本発明の実施例3を示す音声認識処理のフローチャート（全体図）である。

【図10】図9の実施例3の話者分析処理（ステップ804）の詳細フローチャートおよびニューラルネットワークの説明図である。

【図11】本発明の実施例4を示す音声認識処理（全体図）のフローチャートである。

【図12】実施例4で用いるBest-first-searchとスコア推定の説明図である。

【図13】実施例4のテーブル（1010）作成処理のフローチャートである。

【図14】実施例4の図11における認識照合処理（ステップ1002）の詳細フローチャートである。

【図15】本発明の実施例5を示すテーブル（1010）作成処理のフローチャートである。

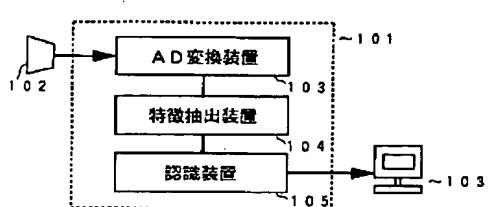
【図16】実施例5の図15における曖昧推論系の決定処理（ステップ1401）の詳細を示すフローチャートである。

【符号の説明】

101…マイクロコンピュータ、102…マイク、103…A/D変換装置、105…認識装置、106…出力装置、420…学習音声サンプル、421…音声データ、422…音声データ、429…音声データ、431…HMMデータ、432…HMMデータ、439…HMMデータ、430…HMM予備データ、511…構成ガ

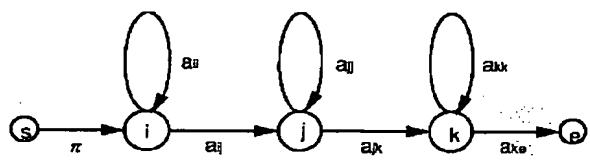
40ウス分布の推論系、512…テーブル値、810…マイク、811…システムデータ、812…出力結果、1010…スコア関数、1210…スコア関数のデータ値、1211…スコア関数の曖昧推論系、1010…テーブル値。

【図1】



[図2]

left-to-right 型 HMM の構成例

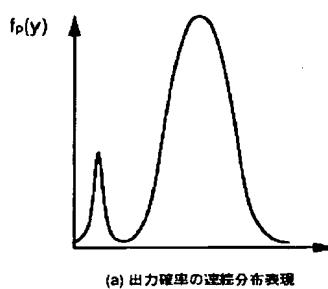


【図3】

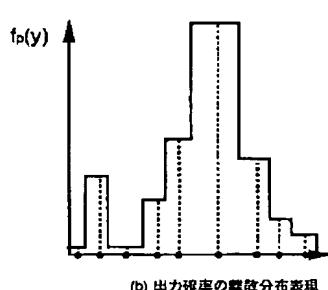
[图 5]

HMMの出力確率の表現形態

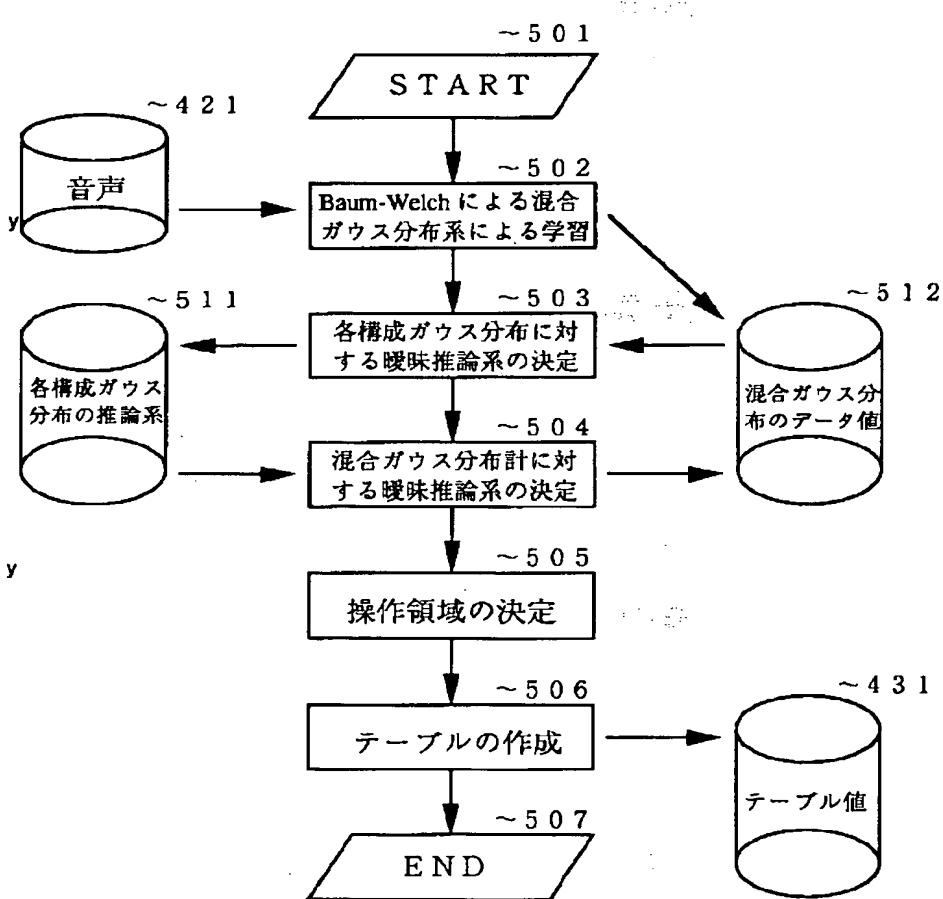
実施例 1 の処理方法（手順 4 1 1）



(a) 出力確率の連続分布表現

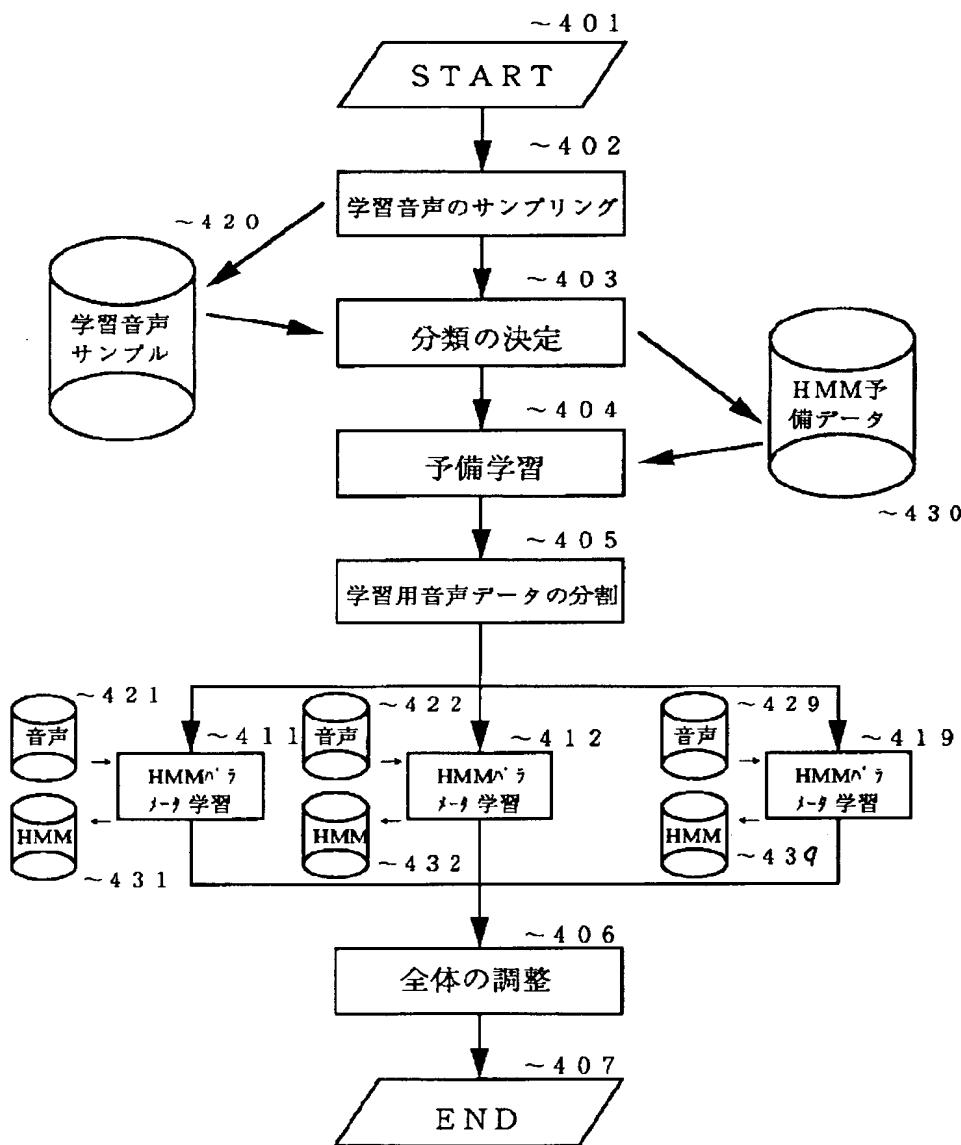


(b) 出力確率の離散分布表現



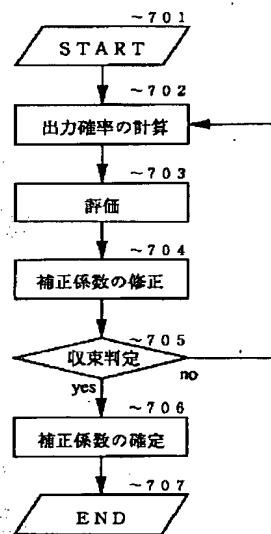
【図4】

実施例1の処理方法（全体図）



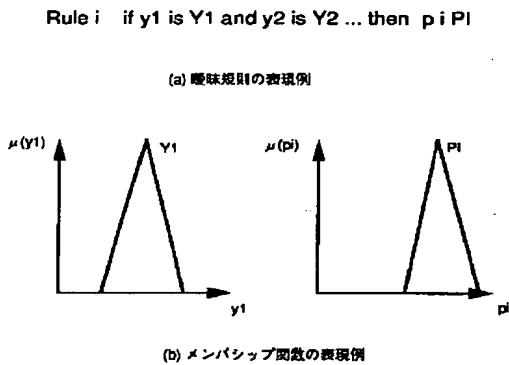
【図7】

実施例1の処理方法（手順406）



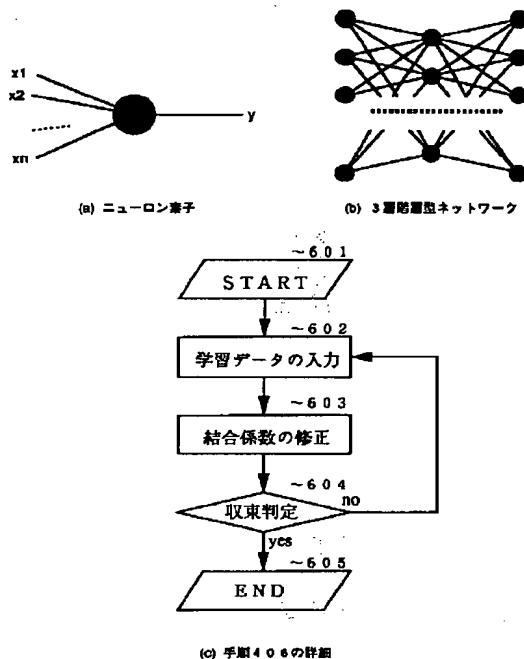
【図6】

曖昧推論による出力確率の推定



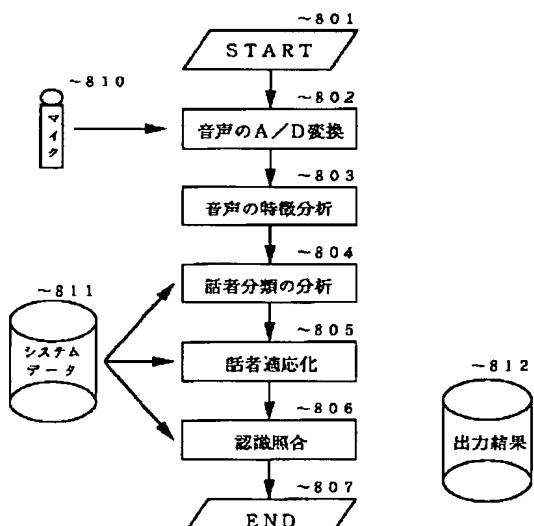
【図8】

実施例2の処理方法



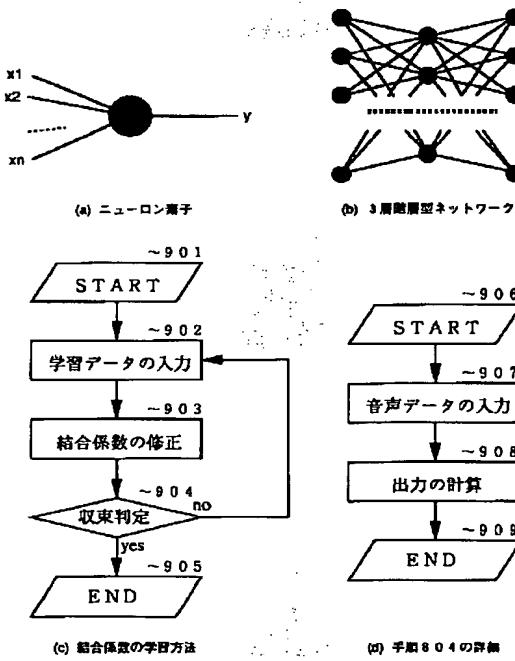
【図9】

実施例3の処理方法（全体図）



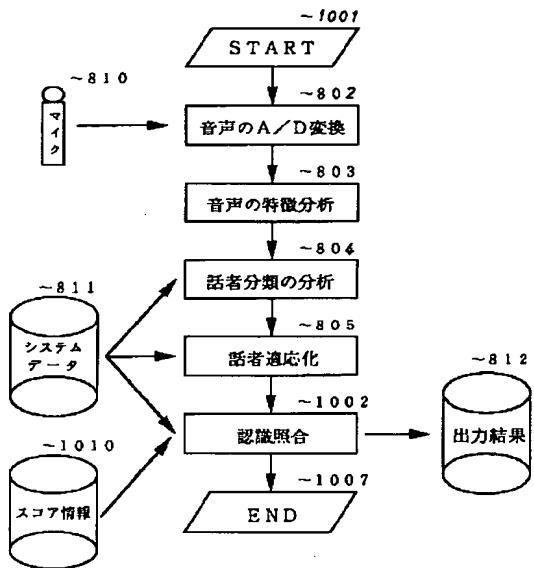
【図10】

実施例3の処理方法（手順804）



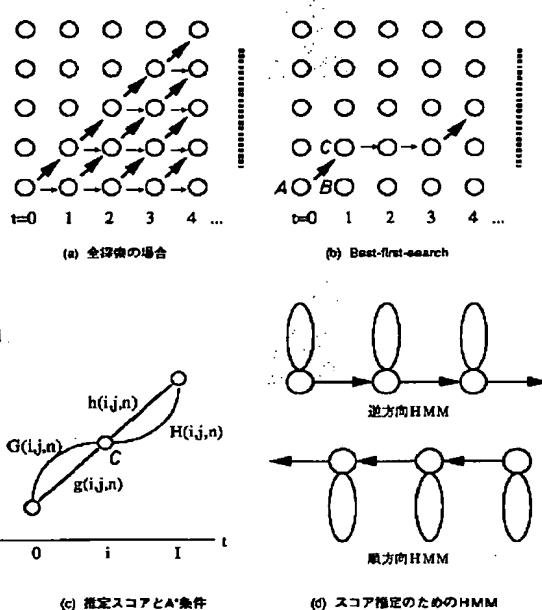
【図11】

実施例4の処理方法（全体図）



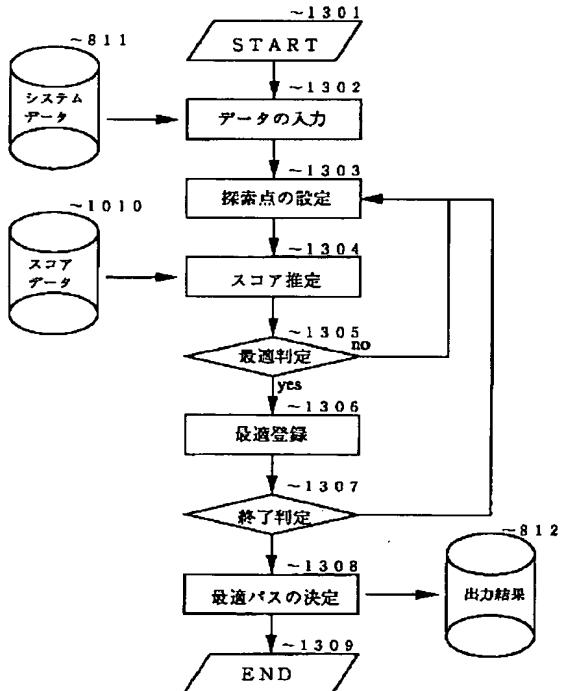
【図12】

Best-first-search とスコア推定



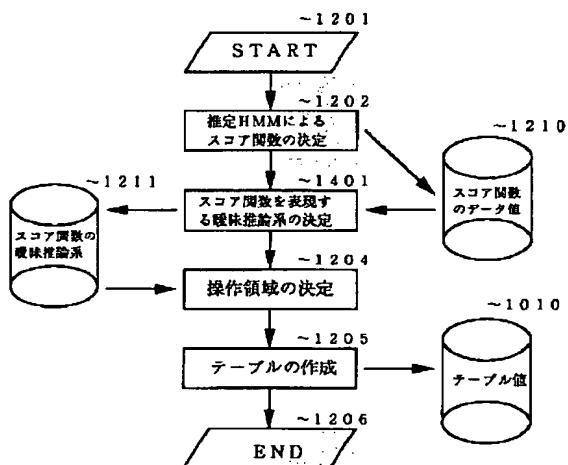
【図14】

実施例4の処理（手順1002）



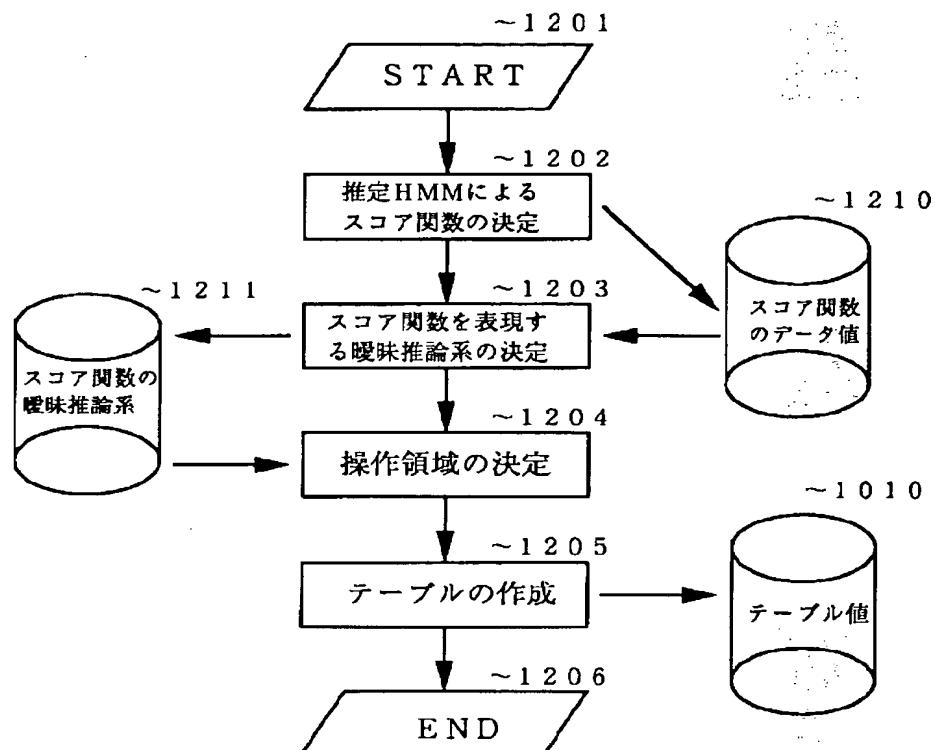
【図15】

実施例5のテーブル1010作成方法



【図13】

実施例4のテーブル1010作成方法



【図16】

